FEBS Letters 583 (2009) 771–776

FEBS Letters

journal homepage: www.FEBSLetters.org

Characterization of the endonuclease SSO2001 from *Sulfolobus solfataricus* P2

Dong Han, Gerhard Krauss *

Department of Biochemistry, Universität Bayreuth, Universitätsstrasse 30, Bayreuth, 95447, Germany

ARTICLE INFO

Article history:
Received 16 December 2008
Revised 16 January 2009
Accepted 16 January 2009
Available online 25 January 2009

Edited by Hans Eklund

Keywords:
CRISPR-associated
Clustered regularly interspaced short palindromic repeats
Nuclease
SSO2001

ABSTRACT

Clustered regularly interspaced short palindromic repeats (CRISPR) and their associated protein genes (*cas* genes) have been suggested to act as an immune system in archaea and bacteria mimicking the eukaryotic RNA interference (RNAi) system. We have investigated the properties of the protein SSO2001 from *Sulfolobus solfataricus* (*Sso*) P2, which is part of the *cas* gene cluster. This study shows that SSO2001 is an endonuclease specifically digesting double-stranded oligonucleotides and preferably cleaving at G:C pairs. Point mutations identify both highly conserved aspartate and glutamate residues as being crucial for the nuclease activity. The catalytic activity shows an optimum at neutral pH and pH 3.

1. Introduction

One only recently recognized defense system in archaea and bacteria, the clustered regularly interspaced short palindromic repeats (CRISPR) system uses regularly spaced repeat sequences and associated genes to fight the invasion from viruses, plasmids, bacteriophage and transposable elements in a manner similar to the eukaryotic RNA interference (RNAi) system [1,2]. The repeat sequences in CRISPR loci have been shown to be derived from foreign DNA. They are organized as direct repetitions of sequences of 26 to 47 bp separated by spacers of 26–72 bp. Sequences and numbers of CRISPR repeats frequently diverge between microbial sequences varying from two to 249 repeats per array. Some microbial species harbor a single CRISPR locus whereas others contain up to 18 loci [3,4].

The gene sequences tightly associated with CRISPR loci have been named *cas* (CRISPR-associated sequences) genes [5]. They are typically organized in small clusters and are associated with specific subtypes of CRISPRs. The number of *cas* genes associated with CRISPR subtypes is also variable, ranging from 6 to more than 20. Specific domains of CAS proteins have been predicted to harbor nuclease, helicase, recombinase, integrase and RNA and/or DNA binding activities [2,6].

Recently, CRISPR from *Streptococcus thermophilus* has been shown to be directly involved in the resistance for infection by bacteriophage. One *cas* protein gene (*cas*7) is directly involved in the process [7]. Another CAS protein has been qualified as a endoribonuclease [8]. The mechanism of acquiring resistance however remains largely unknown.

In the present work, the *cas* gene *sso*2001 from *Sso* P2 was characterized. The gene product SSO2001 belongs to COG2254 identified by computational analysis [1,2,6]. All the homologous protein genes in this cluster are present in archaea and coexist or fuse with *cas*3 protein genes, but none of them has been characterized biochemically to date. SSO2001 was found to be an endonuclease that strongly prefers dsDNA over ssDNA and cleaves preferably at G:C pairs. The optimal activity is at neutral pH and pH 3. Mutational analysis shows that two highly conserved residues, D63 and E92, are both important for enzyme activity.

2. Materials and methods

2.1. Preparation of SSO2001 protein

All cloning steps were performed in *Escherichia coli* strains. The *sso*2001 gene was amplified from *Sso* P2 genomic DNA with the forward primer (5′-ACATATGTTGATCAAGCCTTGCGCTTA) and the reverse primer (5′-ACGAGCTCTAGAGTGGAACCTCCAT). The PCR product was cut out with restriction nucleases *Nde*I and *Sac*I and

Abbreviations: CRISPR, clustered regularly interspaced short palindromic repeats; CAS, CRISPR-associated; COG, cluster of orthologous groups; RNAi, RNA interference; *Sso*, *Sulfolobus solfataricus*; ss(ds), single (double)-stranded; SSO2001-Est, SSO2001-esterase fusion protein; TFK, trifluoremethyl ketone; PAGE, polyacrylamide gel electrophoresis; nt, nucleotide; bp, base pair; MES, 2-(N-morpholino)ethanesulfonic acid; MSH, mercaptoethanol; NDSB, non-detergent sulfobetaines

* Corresponding author. Fax: +49 921 55 2429.
E-mail address: gerhard.krauss@uni-bayreuth.de (G. Krauss).

0014-5793/$34.00
doi:10.1016/j.febslet.2009.01.024

was inserted into the expression vector pET28c (Novagen). The resulting plasmid was transformed into *E. coli* expression strain Rosetta(DE3)pLysS cells (Invitrogen).

The fusion of the esterase from *Alicyclus caldarius* to the SSO2001 protein was achieved by cloning the *sso*2001 gene into the multi-cloning site of the vector pIVEX2.3d-Est2.1 (gift from M. Sprinzl, University Bayreuth, Germany [9]). The fused fragment was cut out with *Nde*I and *BamH*I and was inserted into pET28c to yield a construct in which the *sso*2001 gene contained an N-terminal hexa-histidine tag and was fused C-terminal with the esterase gene.

For expression, inoculated 1 l of LB medium was cultured at 37 °C until OD_{600} was 0.8–1.0. The culture was then cooled down to 30 °C and expression was induced by the addition of 0.1 mM of IPTG. When OD_{600} was >1.5, the cells were harvested by centrifugation and were kept frozen at −70 °C until use.

After cell lysis, the purification of insoluble his-tagged SSO2001 under denaturing conditions was performed on Ni^{2+}-NTA-Sepharose superflow (Qiagen) column. The SSO2001-esterase fusion protein (SSO2001Est) was purified by the Trifluoromethyl ketone (TFK) Sepharose CL-6B affinity column which specifically binds to the esterase. The purified full length proteins were identified by SDS–PAGE, western blotting assay and esterase activity assay [9].

2.2. High-throughput refolding assays

High-throughput assay was performed on a 96-well plate with various buffers as described by Vincentelli et al. [10]. The fractions containing soluble SSO2001 were identified by absorption measurements directly on the plates and were analyzed by SDS–PAGE.

2.3. Preparation of 5′-end labeled oligonucleotide substrates

The 5′-end labeling of DNA or RNA by [γ-^{32}p]-ATP was performed in forward buffer following standard procedures. The labeling reaction contained 2 μM DNA or RNA oligonucleotides, 1 μl [γ-^{32}p]-ATP (10 μci/μl, 2 μM) and 5 U bacteriophage T4 polynucleotide kinase (New England Biolabs).

2.4. Hybridization of oligonucleotides

The mixture of the complementary strands (DNA/DNA, DNA/RNA, or RNA/RNA) was heated in the hybridization buffer (20 mM Tris/AcOH, 10 mM $Mg(OAc)_2$, 50 mM KOAc, 1 mM EDTA, pH 7.9) at 90 °C for 1 min. Then the mixture was cooled down slowly (0.02 °C/s) to 25 °C. Over 95% of the complementary strands were hybridized.

2.5. Nuclease assay

For nuclease activity detection, 10 nM of 5′-labeled DNA or RNA oligonucleotide substrates were incubated with protein in buffer containing 50 mM Tris–Cl, pH 7.5, 10 mM $MgCl_2$ in a total volume of 10 μl for 20 min at 50 °C. The nuclease activity was quenched by addition of denaturing gel loading buffer containing 20 mM EDTA and 0.2% SDS at 4 °C. Samples were then heated to 95 °C for 5 min immediately before electrophoresis on a 20% denaturing PAGE gel. Then the radioactivity was counted by Typ Instant Imager 2024 (Canberra-Packard).

The pH dependence of the nuclease activity was measured using following buffer components: H_3PO_4 for pH 2; glycine for pH 3; NaAc for pH 4; MES for pH 5 and 6; Tris–Cl for pH 7 and 8; CHES for pH 9; and CAPS for pH 10 and 11.

2.6. Preparation of point mutants

Site-directed mutagenesis was carried out using QuikChange kit (stratagene). The mutagenic primers were designed to introduce alanine at the desired sites. Wild-type *sso*2001Est gene cloned in pET28c was used as a template for mutagenesis. Plasmids were purified using Plasmid Miniprep KitII (PeqLab). All mutations were verified by DNA sequencing.

3. Results

3.1. Protein preparation and nuclease detection

In order to characterize the function of the members of the COG2254 cluster, we cloned the gene *sso*2001 and its neighbors, *sso*1998, *sso*1999, and *sso*2002 into the expression vector, pET28c and the co-expression vector, pET-Duet (Novagen). From sequence alignments, SSO1998 was predicted to be an RNA-binding protein, SSO1999 to be a helicase and SSO2001 to be a HD nuclease [6]. There was no functional prediction for SSO2002 (Fig. 1).

Upon expression in *E. coli*, all four proteins could be obtained in insoluble form only. Co-expression of the proteins also did not improve the solubility. Expression level was very high for SSO2001 and therefore renaturation of this protein was tried. Following renaturation of SSO2001, nuclease activity on dsDNA substrates was observed from high-throughput renaturation products [10] in buffers containing 0.8 M arginine at pH 7, 8 and 9 (Fig. 2A, Table 1). However, the refolded SSO2001 remained soluble only for short time and precipitated protein reappeared.

We also tried expression of SSO2001 as fusion with partners that should enhance its solubility. Finally the approach of fusion with esterase from *Alicyclus caldarius* [11] gave a positive result. The use of the esterase as a fusion partner has several advantages. The activity of the esterase present in the full length fusion protein can be monitored in real time by esterase activity assay. The N- and C- termini of esterase are exposed on the protein surface and are not involved in forming the catalytic center of the fused enzyme. The esterase part of the soluble fusion protein could be specifically immobilized on a column containing the esterase inhibitor, TFK. Following washing, the fusion protein could be eluted by addition of the TFK inhibitor [9]. This procedure was successfully applied to

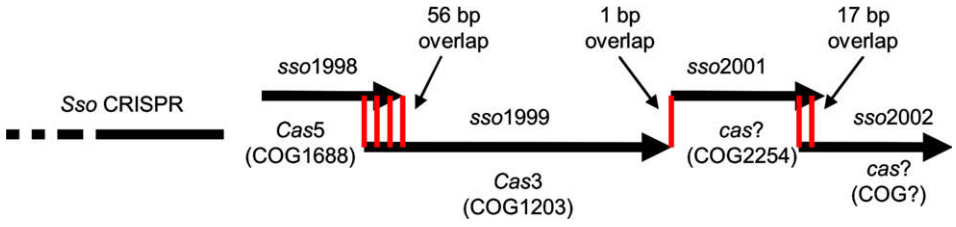

Fig. 1. Location of genes *sso*1998–*sso*2002 close to the *Sso* CRISPR sequence. In sequence alignment, *Sso* P2 protein genes *sso*1998, *sso*1999, *sso*2001 and *sso*2002 are head-to-tail overlapped. They belong to various COGs and *cas* groups except *sso*2001 which matches no *cas* group and *sso*2002 which matches neither COG nor *cas* group. The red bars represent the overlapped regions.

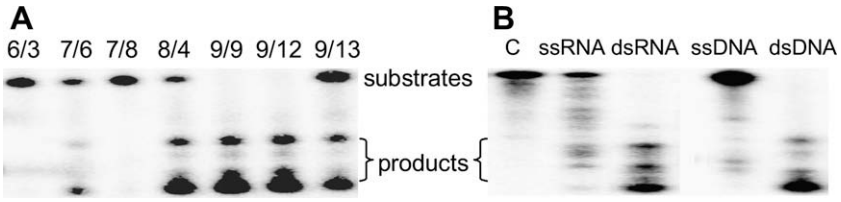

Fig. 2. Nuclease activity detection of SSO2001. Analysis of the reaction products on a 20% denaturing PAGE gel. (A) SSO2001 after high-throughput refolding. The numbering of lanes 6/3∼9/13 represents the well numbers on the refolding plate (see Table 1). (B) SSO2001Est fusion protein with 30 nt (bp) RNA as substrate. Lane C: control sample without protein. (C) SSO2001Est fusion protein with 30 nt (bp) DNA as substrate.

Table 1
Refolding buffers yielding soluble SSO2001.

Well No.	6/3	7/6	7/8	8/4	9/9	9/12	9/13
pH	MES, 6	Tris, 7	Tris, 7	Tris, 8	CHES, 9	CHES, 9	CHES, 9
Refolding buffer components	Glycerol, NDSB256, Arg	PEG4000, NDSB256, Arg	β-MSH, Arg	KCl, β-MSH, NDSB201, Arg	β-MSH, NDSB195, Arg	EDTA, β-MSH, Arg	NDSB195, Arg

Buffer content concentrations: glycerol, 20% (v/v); PEG4000, 0.05% (w/v); NDSB, 100 mM; β-MSH, 10 mM; Arg, 800 mM; EDTA, 1 mM.

the preparation of the fusion protein SSO2001Est, which could be obtained in a stable and soluble form showing nuclease activity.

3.2. Nuclease activity of SSO2001

Various ds and ss DNAs and RNAs were offered as nuclease substrates and the products of the reaction were analyzed on denaturing gels. The gel analysis shows a similar degradation pattern of a dsDNA and a dsRNA molecule of the same sequence into smaller oligomer fragments whereas only minor degradation of ssDNA or ssRNA is observed (Fig. 2B and C). This observation indicates that the native SSO2001Est fusion protein possesses nuclease activity with clear specificity for double-stranded nucleic acids. However, no obvious preference for dsDNA vs dsRNA was found when digestion under different nucleic acid concentrations was followed. Since for dsDNA many more variations were possible for studying the sequence and structure specificity of the nuclease activity we choose dsDNA for the further studies of this work. Comparison of the fusion protein with refolded SSO2001 did not reveal significant differences in the nuclease properties. Thus the further characterization of SSO2001 was carried out with the fusion protein because of its greater stability.

3.3. Optimal condition investigation

Sso is a hyperthermophilic and acidophilic organism and therefore it was necessary to follow the temperature dependence of the nuclease activity of SSO2001Est. Nuclease activity was investigated in the range from 25 °C to 75 °C. The other parameters were fixed. The optimal reaction temperature was found to be 50 °C (Fig. 3A and B).

The pH dependence of nuclease activity was measured in the pH range from 2 to 11. The highest nuclease activity of SSO2001Est was observed in the neutral pH range. Another pH optimum was found at pH 3. The pH dependence is quite unusual with steep increases and decreases on both sides of the optimal pH values (Fig. 3C and D).

The requirement of bivalent metal ions on nuclease activity was investigated. Cleavage of double-stranded substrates by SSO2001-Est was observed in the presence of 10 mM $MgCl_2$, but not within

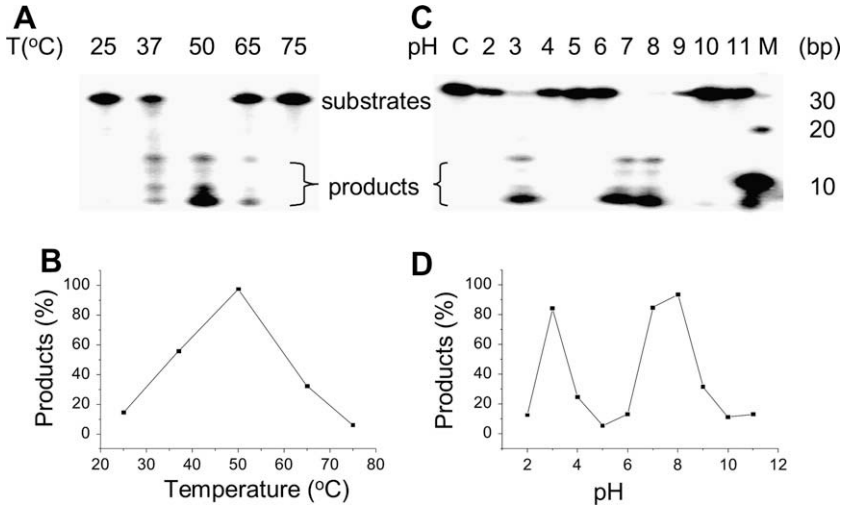

Fig. 3. Temperature dependence (A, B) and pH dependence (C, D) of the nuclease activity of SSO2001Est. 5′-end labeled 30 bp dsDNA (see Table 2) was used as a substrate. Samples were analyzed by 20% denaturing PAGE. In (C), lane C represents the control without protein, and M is the 10 bp DNA marker.

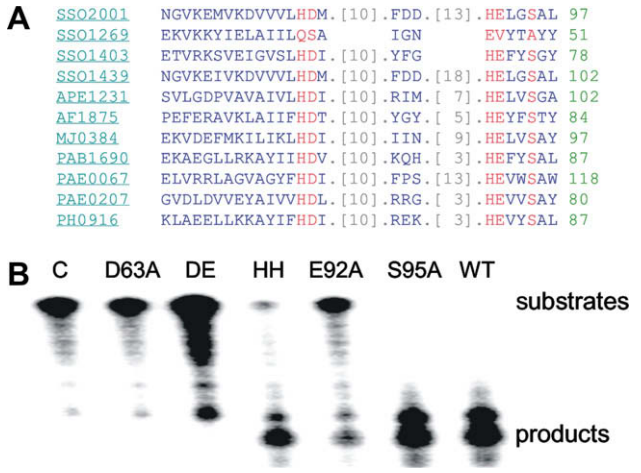

Fig. 4. Mutational analysis of SSO2001Est. (A) Multi-sequence alignment of SSO2001 homologues by BLASTP. The highlighted red letters represent the conserved residues. All putative proteins are from Archaea and possess the highly conserved HD motifs, HE motifs and serine residues. Single letters in sequence represent amino acid residues. (B) Nuclease activity of point mutants of SSO2001Est. The substrate was a 5′-end labeled 30 bp dsDNA (see Table 2). lane C represents the control without protein. The numbers represent the amino acid residue positions in original SSO2001 sequence. WT represents the wild-type of SSO2001Est.

Ca^{2+}, Mn^{2+}, Ni^{2+}, or Zn^{2+} ions. EDTA inhibited the nuclease activity by chelating Mg^{2+} ions.

3.4. Protein active site determination by point mutation

To verify the intrinsic nuclease activity of SSO2001 and to exclude contaminating nuclease activity, it was necessary to create inactivating point mutants of SSO2001. The multi-alignment of SSO2001 sequence by BlastP program on NCBI website showed that all proteins hit by the program were hypothetical and were from Archaea. All of them belong to COG2254 and were predicted as core proteins in CRISPR system [1], however none of them has been characterized to date. Three highly conserved motifs could be identified by the sequence alignment, a HD motif, a HE motif and a conserved Serine residue (Fig. 4A). The presence of a highly conserved HD motif (H62, D63) on these proteins had suggested that they belonged to the HD hydrolase superfamily [12]. The HE motif (H91, E92) had not been mentioned as a signature sequence for nucleases, neither was the conserved serine residue (S95).

To determine the catalytically important residues of SSO2001, alanine point mutants in the SSO2001 sequence were created on the SSO2001Est fusion protein. The point mutants are D63A, E92A and S95A, and the double mutants are H62A-H91A and D63A-E92A (the numbers represent the positions of the amino acid residues in the original SSO2001 sequence), respectively. The degradation patterns indicate that the D63 residue is most important for nuclease activity. The activity is nearly completely abolished in the D63A mutant. The activity of the E92A mutant is also diminished, however some residual activity remains. The mutation of S95 does not disturb the nuclease activity at all. The nuclease activity of the double mutant D63A-E92A is completely inhibited which shows that these two residues are crucial for nuclease activity. The double mutant H62A-H91A shows only slightly decreased nuclease activity (Fig. 4B).

3.5. Substrate specificity of nuclease activity

An important point to be investigated was the structure specificity and the sequence specificity of the nuclease activity. Various DNA oligonucleotides including paired double-strands, pseudo fork-strands and double-strands with sticky ends were prepared as substrates for nuclease detection (see the schematic map of the substrate structures in Fig. 5B). The cleavage reaction was quenched after 10 min to obtain partially degraded products. The degradation of 5′-end labeled upper or bottom strands in dsDNA oligonucleotides and double strands with sticky ends (Fig. 5A, lanes 1 to 8) indicates an endonuclease activity. The cleavage stopped at approximately 4 base pairs from the 5′-end implying that dsDNA strands shorter than 4 base pairs are not optimal substrates for nuclease activity (Fig. 5A, lanes 7 and 8). The cleavage of pseudo fork-strands (Fig. 5A, lanes 9, 10, 11 and 12) shows the double strand preference of the enzyme activity. The sequence (Table 2) analysis shows that the main products arise from enzyme cleavage at G:C pairs (pointed by black arrows in Fig. 5A) suggesting that the enzyme preferentially cleaves the dsDNA at G:C pairs. No structure specificity of the enzyme activity was observed from this experiment. Further experiments with CRISPR repeat or spacer sequences as substrates did not show any preference of SSO2001 for these sequences (data not shown).

4. Discussion

In our laboratory, over ten putative *cas* protein genes from *Sulfolobus* had been tried to be expressed in soluble active form in various *E. coli* expression systems. Soluble expression of most of the

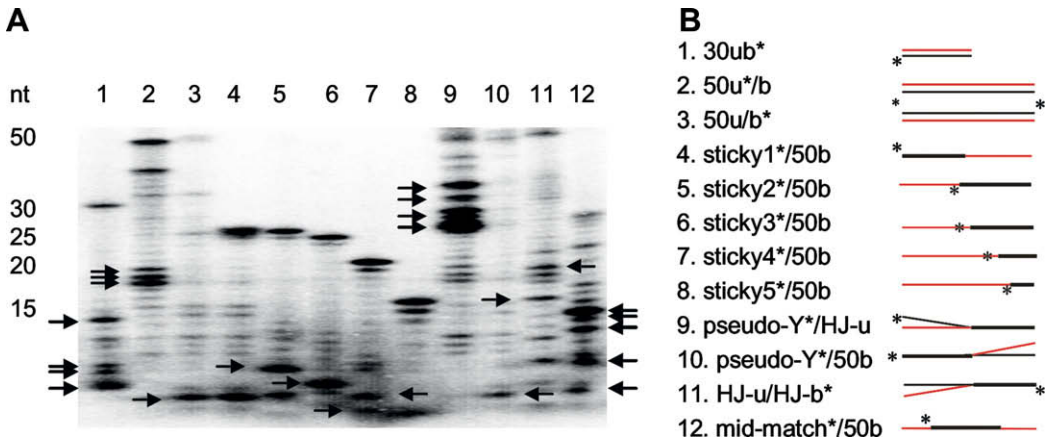

Fig. 5. Digestion of various 5′-end radioactively labeled oligonucleotide substrates by SSO2001Est. Samples were analyzed on a 20% sequencing gel. Arrows in (A) indicate cleavage products with G:C pairs at the cleavage point. In (B), the structures of the oligonucleotides are indicated schematically (for sequences, see Table 2) and the radioactively labeled 5′-ends are marked with black stars.

Table 2
Sequences of DNA substrates.

Name	Sequence
1. 30ub*	5′-CCTCTTCTTCTGTGCACTCTTCTTCTCCCC –3′ 3′-GGAGAAGAAGACACGTGAGAAGAAGAGGGG*-5′
2. 50u*/b	5′-*ACAGCTATGACCGAATTCCTGGGGAGAAGAAGAGTGCACAGAAGAAGAGG-3′ 3′-TGTCGATACTGGCTTAAGGACCCCTCTTCTTCTCACGTGTCTTCTTCTCC-5′
3. 50u/b*	5′-ACAGCTATGACCGAATTCCTGGGGAGAAGAAGAGTGCACAGAAGAAGAGG-3′ 3′-TGTCGATACTGGCTTAAGGACCCCTCTTCTTCTCACGTGTCTTCTTCTCC*-5
4. sticky1*/50b	5′-*ACAGCTATGACCGAATTCCTGGGGA-3′ 3′-TGTCGATACTGGCTTAAGGACCCCTCTTCTTCTCACGTGTCTTCTTCTCC-5′
5. sticky2*/50b	5′-*GAAGAAGAGTGCACAGAAGAAGAGG-3′ 3′-TGTCGATACTGGCTTAAGGACCCCTCTTCTTCTCACGTGTCTTCTTCTCC-5′
6. sticky3*/50b	5′-*AAGAAGAGTGCACAGAAGAAGAGG-3′ 3′-TGTCGATACTGGCTTAAGGACCCCTCTTCTTCTCACGTGTCTTCTTCTCC-5′
7. sticky4*/50b	5′-*AGAGTGCACAGAAGAAGAGG-3′ 3′-TGTCGATACTGGCTTAAGGACCCCTCTTCTTCTCACGTGTCTTCTTCTCC-5′
8. sticky5*/50b	5′-*GCACAGAAGAAGAGG-3′ 3′-TGTCGATACTGGCTTAAGGACCCCTCTTCTTCTCACGTGTCTTCTTCTCC-5′
9. pseudo-Y*/HJ-u	5′-*ACAGCTATGACCGAATTCCTGGGGACCTGGTCGACCTGCAGGCATGCAAG-3′ 3′-AAGGACGAGCTCCCGCGGTCCCACCGGACCAGCTGGACGTCCGTACGTTC-5′
10. pseudo-Y*/50b	5′-*ACAGCTATGACCGAATTCCTGGGGACCTGGTCGACCTGCAGGCATGCAAG-3′ 3′-TGTCGATACTGGCTTAAGGACCCCTCTTCTTCTCACGTGTCTTCTTCTCC-5′
11. HJ-u/HJ-b*	5′-CTTGCATGCCTGCAGGTCGACCAGGCCACCCTGGCGCCCTCGAGCAGGAA-3′ 3′-GGAGAAGAAGACACGTGAGAAGAAGGGTGGGACCGCGGGAGCTCGTCCTT*-5′
12. mid-match*/50b	5′-*GAATTCCTGGGGAGAAGAAGAGTGCA-3′ 3′-TGTCGATACTGGCTTAAGGACCCCTCTTCTTCTCACGTGTCTTCTTCTCC-5′

Black stars represent radioactively labeled 5′-end of oligonucleotides; red letters, the paired bases in sequences.

proteins failed, including *sso*2001. The ORFs of *sso*1998–*sso*2002 are arranged in *Sso* in an overlapping way (Fig. 1) which not only implies a functional relation but also suggests co-expression of these genes in *Sso* [13]. However, the co-expression of these genes in *E. coli* failed in our lab.

Although the activity profile of refolded SSO2001 was the same as that of SSO2001Est, renatured SSO2001 showed a strong tendency to aggregate suggesting that SSO2001 might require other protein partners to stay in a soluble form. The fact that the folding of SSO2001 is facilitated by the presence of the well soluble esterase from *Alicyclobacillus acidocaldorius* indicates that the esterase performs a chaperon-like function. Probably *Sso*-specific chaperons are required for expression *in vivo*. A few *Sso* chaperons have been described [14–16], however these proteins were not available for the present work.

The cleavage reaction is highly dependent on the pH of the buffer. The optimal pH of the nuclease activity of SSO2001Est fusion protein was in the range of neutral pH (7–8). However, an approximately 10% lower activity was found at pH 3. Only few examples of two well separated pH optima have been described. One example is sialidase which exists in different states at corresponding pHs [17].

The optimal temperature of SSO2001Est fusion protein was found to be 50 °C. For an enzyme from a thermophilic organism with an optimal growth temperature of around 75 °C, one would expect a higher temperature optimum. The lower temperature optimum might be due to a balance of the thermostability between the thermophilic enzyme SSO2001 and the esterase which optimal temperature is 55 °C. Since the esterase is important for the solubility of SSO2001, this suggests that the temperature optimum of the fusion protein is not the real optimum of SSO2001 *in vivo*.

The cleavage occurs more efficiently in the vicinity of G:C base pairs. A similar base preference has been reported for nucleases from *Streptomyces antibioticus*, *Neurospora crassa* and *Chlamydomonas reinhardtii* that show a preference for dG bases in dsDNA or ssDNA oligonucleotide [18–20], although the majority of nucleases reported so far are base non-specific.

Homology search predicted SSO2001 to be a HD-like nuclease [2,6]. In HD-superfamily proteins, histidine and aspartate residues cooperate with metal ions and are directly involved in the catalytic reaction [12,21]. The alignment of SSO2001 with its homologues identified, in addition to the HD sequence, a highly conserved HE motif and a conserved serine residue. The HE motif is not found in HD-superfamily enzymes; neither is the conserved S95. The results from our point mutation studies indicate that the highly conserved aspartate and glutamate residues in HD and HE motifs are essential for the activity whereas the histidine appears to be not essential. Possibly both participate in magnesium ion binding and catalytic action. Therefore, our data suggest that SSO2001 does not belong to the HD-superfamily of nuclease.

A recent work has shown that COG2254 proteins together with CAS3 protein (COG1203) specifically cleave transcripts of the CRISPR repeat region into 70∼100 nt pre-prokaryotic siRNA [2]. In our work we could not find a preference of SSO2001 for dsRNA as compared to dsDNA. Furthermore, dsDNA harboring the CRISPR repeat and spacer sequences is not preferentially cleaved as compared to CRISPR unrelated sequences. This lack of specificity might be due to the lack of partner proteins, possibly SSO1999 (CAS3 protein), or unknown specific CRISPR recognition proteins. A possible function of the nuclease activity of SSO2001 might be the excision of CRISPR sequences from dsDNA or RNA-DNA hybrids at some points in the process of preventing invasion of foreign DNA.

Acknowledgements

We thank Prof. Mathias Sprinzl and Dr. Yiwei Huang for providing pIVEX2.3d-Est2.1 vector and TFK resin and Dr. Georg Lipps for kindly advice. This work was supported by grant Kr 704/ 12-2 from the Deutsche Forschungsgemeinschaft to G.K.

References

[1] Haft, D.H., Selengut, J., Mongodin, E.F. and Nelson, K.E. (2005) A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. PLoS. Comput. Biol. 1, 474–483.
[2] Makarova, K.S., Grishin, N.V., Shabalina, S.A., Wolf, Y.I. and Koonin, E.V. (2006) A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. Biol. Direct. 1, 7.
[3] Godde, J.S. and Bickerton, A. (2006) The repetitive DNA elements called CRISPRs and their associated genes: evidence of horizontal transfer among prokaryotes. J. Mol. Evol. 62, 718–729.

[4] Mojica, F.J., ez-Villasenor, C., Garcia-Martinez, J. and Soria, E. (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. J. Mol. Evol. 60, 174–182.
[5] Jansen, R., Embden, J.D., Gaastra, W. and Schouls, L.M. (2002) Identification of genes that are associated with DNA repeats in prokaryotes. Mol. Microbiol. 43, 1565–1575.
[6] Makarova, K.S., Aravind, L., Grishin, N.V., Rogozin, I.B. and Koonin, E.V. (2002) A DNA repair system specific for thermophilic Archaea and bacteria predicted by genomic context analysis. Nucleic Acids Res. 30, 482–496.
[7] Marx, J. (2007) Microbiology. New bacterial defense against phage invaders identified. Science 315, 1650–1651.
[8] Beloglazova, N., Brown, G., Zimmerman, M.D., Proudfoot, M., Makarova, K.S., Kudritska, M., Kochinyan, S., Wang, S., Chruszcz, M., Minor, W., Koonin, E.V., Edwards, A.M., Savchenko, A. and Yakunin, A.F. (2008) A novel family of sequence-specific endoribonucleases associated with the clustered regularly interspaced short palindromic repeats. J. Biol. Chem. 283, 20361–20371.
[9] Huang, Y., Humenik, M. and Sprinzl, M. (2007) Esterase 2 from *Alicyclobacillus acidocaldarius* as a reporter and affinity tag for expression and single step purification of polypeptides. Protein Express Purif. 54, 94–100.
[10] Vincentelli, R., Canaan, S., Campanacci, V., Valencia, C., Maurin, D., Frassinetti, F., Scappucini-Calvo, L., Bourne, Y., Cambillau, C. and Bignon, C. (2004) High-throughput automated refolding screening of inclusion bodies. Protein Sci. 13, 2782–2792.
[11] Agafonov, D.E., Rabe, K.S., Grote, M., Huang, Y. and Sprinzl, M. (2005) The esterase from *Alicyclobacillus acidocaldarius* as a reporter enzyme and affinity tag for protein biosynthesis. FEBS Lett. 579, 2082–2086.
[12] Aravind, L. and Koonin, E.V. (1998) The HD domain defines a new superfamily of metal-dependent phosphohydrolases. Trends Biochem. Sci. 23, 469–472.
[13] Johnson, Z.I. and Chisholm, S.W. (2004) Properties of overlapping genes are conserved across microbial genomes. Genome Res. 14, 2268–2272.
[14] Guagliardi, A., Cerchia, L., Bartolucci, S. and Rossi, M. (1994) The chaperonin from the archaeon *Sulfolobus solfataricus* promotes correct refolding and prevents thermal denaturation in vitro. Protein Sci. 3, 1436–1443.
[15] Ruggero, D., Ciammaruconi, A. and Londei, P. (1998) The chaperonin of the archaeon *Sulfolobus solfataricus* is an RNA-binding protein that participates in ribosomal RNA processing. EMBO J. 17, 3471–3477.
[16] Guagliardi, A., Mancusi, L. and Rossi, M. (2004) Reversion of protein aggregation mediated by Sso7d in cell extracts of *Sulfolobus solfataricus*. Biochem. J. 381, 249–255.
[17] Venerando, B., Fiorilli, A., Croci, G., Tringali, C., Goi, G., Mazzanti, L., Curatola, G., Segalini, G., Massaccesi, L., Lombardo, A. and Tettamanti, G. (2003) Acidic and neutral sialidase in the erythrocytes of patients with type 2 diabetes: an answer to comments by Richard et al.. Blood 101, 2071.
[18] Burton, W.G., Roberts, R.J., Myers, P.A. and Sager, R. (1977) A site-specific single-strand endonuclease from the eukaryote Chlamydomonas. Proc. Natl. Acad. Sci. USA 74, 2687–2691.
[19] Linn, S. and Lehman, I.R. (1965) An endonuclease from neurospora crissal specific for polynucleotides lacking an ordered structure. II. Studies of enzyme specificity. J. Biol. Chem. 240, 1294–1304.
[20] Cal, S., Nicieza, R.G., Connolly, B.A. and Sanchez, J. (1996) Interaction of the periplasmic dG-selective *Streptomyces antibioticus* nuclease with oligodeoxynucleotide substrates. Biochemistry 35, 10828–10836.
[21] Yakunin, A.F., Proudfoot, M., Kuznetsova, E., Savchenko, A., Brown, G., Arrowsmith, C.H. and Edwards, A.M. (2004) The HD domain of the Escherichia coli tRNA nucleotidyltransferase has 2′,3′-cyclic phosphodiesterase, 2′-nucleotidase, and phosphatase activities. J. Biol. Chem. 279, 36819–36827.